

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : G10L 5/06	A1	(11) International Publication Number: WO 99/16051 (43) International Publication Date: 1 April 1999 (01.04.99)
(21) International Application Number: PCT/IB98/01717 (22) International Filing Date: 24 September 1998 (24.09.98) (30) Priority Data: 60/059,895 24 September 1997 (24.09.97) US (71) Applicant: LERNOUT & HAUSPIE SPEECH PRODUCTS N.V [BE/BE]; Sint-Krispijnstraat 7, B-8900 Ieper (BE). (72) Inventors: SCHOOF, Koen; Milsestraat 65, B-3050 Haasrode (BE). GALLOPYN, Guido; Astridstraat 7, B-9620 Zottegem (BE). (74) Agent: FROUD, Clive; Elkington and Fife, Prospect House, 8 Pembroke Road, Sevenoaks, Kent TN13 1XR (GB).		(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
(54) Title: APPARATUS AND METHOD FOR DISTINGUISHING SIMILAR-SOUNDING UTTERANCES IN SPEECH RECOGNITION		
(57) Abstract <p>An apparatus and method use speaker-specified hints to establish conditions for a speech recognition system to select a recognition result for a previously provided utterance from among various possible homophones. The hints may characterize the utterance by a linguistic property, such as an orthographic, morphological, or semantic property.</p> <div style="text-align: center;"> <pre> graph TD A([PROVIDED UTTERANCE]) --> B[IDENTIFY SPEAKER-SPECIFIED HINT 21] B --> C[ASSOCIATE HINT WITH PROVIDED UTTERANCE 22] C --> D[ESTABLISH CONDITION FOR DISTINGUISHING UTTERANCE 23] D --> E[PROVIDE LIST OF ALTERNATIVE RECOGNITION POSSIBILITIES 24] E --> F[FILTER LIST BASED ON CONDITION 25] F --> G[SELECT RECOGNITION RESULT SATISFYING CONDITION 26] G --> H([RECOGNITION RESULT]) </pre> </div>		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

Apparatus and Method for Distinguishing Similar-Sounding Utterances in Speech Recognition

Technical Field

The present invention relates to speech recognition systems, and more particularly, to dictation systems.

Background

One of the most difficult problems for speech recognition systems is how to choose the correct alternative from a list of words that are pronounced similarly or identically, but spelled differently. Such similar sounding words are known as homophones. For example, in English, "read, red" or "send, sent" are such homophones, as are the French words "parle, parles, parlent". Humans select the correct homophone by considering the context in which a given word appears, or by understanding the content of the text. This technique is, however, not yet feasible for computer systems.

In current computer-based speech recognition systems, one alternative is selected. If that alternative is incorrect, the user may then go and correct the selection, for example, by choosing another alternative from a list of similar-sounding words. This method has the disadvantage that the user must spot the mistake in the recognized text, and then correct it. This takes extra time, breaks the flow of dictation, and carries the risk that some errors may be overlooked.

Summary of the Invention

In a preferred embodiment, the invention provides a method of utilizing a speech recognizer to distinguish a provided utterance from one or more similar-sounding utterances when a speaker-specified hint is provided. The method includes:

- a. identifying the hint and associating with it the provided utterance;

- b. using the hint to establish a condition for distinguishing the provided utterance; and
- c. selecting a recognition result, derived in conjunction with operation of the speech recognizer, that satisfies the condition. In further
5 embodiments, step (c) includes:
 - (i) providing a list of alternative recognition possibilities and
 - (ii) filtering the list based upon the condition.

The speech recognizer may be used to provide entries in the list of alternative recognition possibilities. Alternatively, or in addition, a dictionary
10 may be used to provide entries in the list. The hint may be a linguistic property characterizing the provided utterance or, in the same or an alternative embodiment, the hint may make reference to the context of previous dictation to characterize the provided utterance. Where the hint is a linguistic property, it may be an orthographic, morphological, or semantic property of the provided
15 utterance. In a simple embodiment, for example, when the hint is an orthographic property, it may be a fractional spelling of the provided utterance. (As used in this description and in the following claims, the term "fractional spelling" means providing the spelling for a portion of the word, wherein the portion need not, but might possibly, include the beginning of the word, and
20 does not include the whole word).

Alternatively, the hint may provide some other desired criterion for selecting the provided utterance. The utterance may be a word, or alternatively, a phrase. For the purposes of this description and the following claims, a "hint" excludes a complete spelling of a word or phrase. Also, alternatively, a plurality
25 of hints may be utilized. In such an embodiment, the method includes:

- (a) identifying the hints and associating with them the provided utterance;
- (b) using the hints to establish conditions for distinguishing the provided utterance; and
- 30 (c) selecting a recognition result, derived in conjunction with operation of the speech recognizer, that satisfies the conditions.

A related embodiment provides an improved speech recognition system, of the type providing a text output in response to a spoken input. The improvement renders the system capable of distinguishing a provided utterance from one or more similar-sounding utterances when a speaker-identified hint is provided. The improvement includes:

- a. a hint recognizer which identifies the hint and associates with it the provided utterance;
- b. a condition specifier, coupled to the hint recognizer, which uses the hint to establish a condition for distinguishing the provided utterance; and
- 10 c. a result selector, coupled to the condition specifier, which selects a recognition result that satisfies the condition.

In a further embodiment, the result selector includes a filter operative on a list of alternative recognition possibilities. In a still further embodiment, the improvement also includes a dictionary, to which the result selector is coupled,

15 to provide entries in the list of alternative recognition possibilities. As before, the hint may be a linguistic (e.g. orthographic, morphological, or semantic) property characterizing the provided utterance, or may make reference to the context of previous dictation to characterize the provided utterance, or may provide some other desired criterion for selecting the provided utterance.

20 In yet a further embodiment, there is provided a system for utilizing a plurality of hints. The system of this embodiment includes:

- (a) a hint recognizer for identifying the hints and associating with them the provided utterance;
- (b) a condition specifier, coupled to the hint recognizer, for using the
- 25 hints to establish conditions for distinguishing the provided utterance; and
- (c) a result selector, coupled to the condition specifier, for selecting a recognition result that satisfies the conditions.

Another embodiment includes an improved speech recognition system wherein the improvement renders the system capable of distinguishing a

30 provided utterance from one or more similar-sounding utterances when a speaker-identified hint is provided. The improvement includes:

- (a) means for identifying the hint and associating with it the provided utterance;
- (b) means for using the hint to establish a condition for distinguishing the provided utterance; and
- 5 (c) means for selecting a recognition result, derived in conjunction with operation of the speech recognizer, that satisfies the condition.

Brief Description of the Drawings

The foregoing aspects of the invention will be more readily understood by
10 reference to the following detailed description taken with the accompanying drawings in which:

Fig. 1 is a block diagram of a system to which the present invention is applicable.

Fig. 2 is a logical flowchart of a method in accordance with a preferred
15 embodiment of the invention.

Detailed Description of Specific Embodiments

A preferred embodiment of the present invention provides a solution to the homophone problem by allowing the user to give hints about the correct
20 spelling of a word. Such user-provided hints may be given in an intuitive manner, while the dictation is going on, without requiring that the word be spelled in its entirety. In the example of "send, sent" the hint may be that the user dictates: "sent with-a-t" or "send with-a-d". A preferred embodiment may also be used for Asian languages to give hints about which character (KANJI, for
25 example) to select for a word. Such a preferred embodiment may be used in virtually any language, particularly in languages having many homophonic words such as French and Chinese.

Thus, as shown in Fig. 1, a preferred embodiment includes a speech recognition engine 1 such as is well-known in the art. For example, the speech
30 recognition engine 1 may be a large vocabulary continuous speech recognition engine such as that used in VoiceXpress™ manufactured by Lernout & Hauspie Speech Products N.V., located in Burlington, MA. Further information on the

design of a speech recognition system is provided, for example, in Rabiner and Juang, *Fundamentals of Speech Recognition*, Prentice Hall 1993, which is hereby incorporated herein by reference.

In communication with the speech recognition engine 1 is a hint identifier
5 and homophone selector 2 which identifies a speaker-specified hint received by the speech recognition engine 1, step 21 in Fig. 2. Such a hint is essentially a command to the speech recognition engine 1 rather than dictated text.

Distinguishing spoken commands from dictated text is known in the art and described, for example, in U.S. Patent 5,794,196, issued to Yegnanarayanan et al.,
10 incorporated herein by reference, which describes recognizing isolated word commands embedded in a stream of continuously dictated text, and utilized in the VoiceXpress™ product which distinguishes continuously spoken natural language commands from continuously dictated text. Accordingly, the specific details of distinguishing a hint from dictated text are not relevant to the present
15 invention.

In step 22 of Fig. 2, the hint is associated with a previously provided input utterance. Based upon the hint, the hint identifier and homophone selector 2 establishes a condition for distinguishing the provided utterance, step 23 of Fig. 2. In step 24 a list may be provided of alternative recognition possibilities for the
20 provided utterance. Typically, a dictionary 3 with linked entries will be available to the speech recognition engine 1 to provide such alternative recognition possibilities. In step 25, the list of alternative recognition possibilities is then filtered based on the condition established from the hint. Lastly, the hint identifier and homophone selector 2 will select, in step 26, the recognition result
25 which satisfies the condition.

In a preferred embodiment, system commands may be provided such as "spelled with X", "spelled with an X", "with an X", "with X", "without X", or the equivalent. In such a dictation system, X may be one or more letters of the alphabet in the language of the system, e.g., "d", "t", "double-l", "dt" (typical for
30 Dutch). Alternatively, X may be a description or hint of a KANJI character, for example, related to the number of strokes of the character, or some other feature used to distinguish between KANJI characters. Other orthographic hints are also

within the scope of a preferred embodiment. For example, a hint may include a designation of the alphabet in which the utterance is rendered, *e.g.*, in Japanese, hiragana or katakana.

As an alternative, or in addition, system commands may supply a hint
5 about a morphological property, *e.g.*, singular, plural, or past tense.
Furthermore, a hint may be some other linguistic property characterizing the provided utterance—for example, a semantic property such as “the color” when uttered after “red”. Such a hint also may refer to the context of previously dictated text to characterize a word, *e.g.*, “as I used it before”. These hints can be
10 given by the user without disrupting the flow of the dictation.

When a “Y spelled with X” command is recognized, the system looks up Y and considers various alternatives—homophones or alternative recognition results. These alternatives can be retrieved from a dictionary which contains links between all words that sound alike (sometimes referred to below and in the
15 claims as a “linked dictionary”), or, from an alternatives list in the recognition result of the word Y. The alternatives are then filtered based on whether they fulfill the command criterion “X” (or not, in a case of “without X”). From the remaining alternatives, one is chosen, *e.g.*, the alternative with the highest occurrence probability.

20 The alternative filtering may be achieved in various manners. In one approach, all alternatives which lack the characters described in X are removed. Some pre-processing may be done, for example, “double L” would be replaced by “LL”, then all alternatives not containing “LL” are removed. Another filtering approach exploits the fact that many hints are related to verb endings
25 (“sent with a t”). Accordingly, the system may check whether the last letter(s) of the verb correspond to X. In this manner, X can be restrained to commonly confusable verb endings (*e.g.*, d, t for English; e, s, es, t, ent for French). In another filtering approach, identifiers in a dictionary may be utilized to show to which letter a hint applies, if present (an index to a start position in the word
30 string would suffice). For example, to differentiate KANJI characters, the hint may be stored in the dictionary entry for a word, such as in a field indicating the number of strokes in the character.

As an alternative to generating a list of alternatives and then filtering for the condition established by the hint, it is within the scope of the present invention to use the condition in connection with a linked dictionary to produce directly a single recognition result satisfying the condition.

5 A preferred embodiment also has language model and grammar implications. In speech recognition, a word or a command can only be recognized if it is part of a grammar of a language model. This also applies to the hints as used in a preferred embodiment. Different options are possible to add hints to a language model. For example, the hint phrase "spelled with" may
10 be modeled in the same way as a "capitalize that" command. That is, the hint can occur at any point in the dictation, after any word. This can be modeled by giving the hint a unigram occurrence probability. The value of the probability should be in line with the probability assigned to other commands such as "capitalize that". Alternatively, "spelled with" may be constrained to occurring
15 only after certain classes of confusable words; for example, only after verbs.

In a hint such as "Y spelled with X", the X can also be modeled in different ways. For example, X may be part of the language model and treated in the same way as any other word, e.g., as a unigram, or bigram. Thus, a probability can be computed for the transitions "with X", "with a X", and "with
20 an X" in the same way as with other recognition probabilities. Alternatively, X may be treated as a limited domain spelling grammar, which is entered when the system recognizes the phrase "spelled with". The grammar would incorporate all commonly given hints. Similarly, the recognition system may switch modes to a spelling grammar to recognize X. Or, the phrase "spelled with X" may be
25 treated as a separate grammar. This grammar may be entered through normal dictation, or it may be activated when displaying an alternative list as a separate window.

Although preferred embodiments have been described above with respect to the use of a single hint, it is within the scope of the present invention to
30 provide a plurality of hints. In French, for example, one might usefully indicate that a verb is feminine and singular, thereby providing two hints. Although the present invention is particularly applicable to continuous dictation systems, it is

- also applicable to discrete dictation systems. Furthermore, while the invention may be employed for hint-giving during dictation, it may also be applied as a correction mechanism for text that has already been dictated. For example, hints may be used to select a recognition result from a displayed list of alternative
- 5 recognition possibilities appearing in a window separate from the text. In this way a "Y spelled with X" command is not embedded in normal dictation mode, but provides a novel way to select alternatives in an alternatives list.
- Alternatively, when in the correction mode, a hint may cause the immediate selection of a recognition result, without the display of an alternatives list, in the
- 10 manner described above for dictation. In such an embodiment, a hint need not necessarily be identified as such, since the system is already in a correction mode and may be configured to act on the hint directly without needing an identification step or a hint recognizer.

What is claimed is:

1. A method of utilizing a speech recognizer to distinguish a provided utterance from one or more similar-sounding utterances when a speaker-specified hint is provided, the method comprising:
 - 5 (a) identifying the hint and associating with it the provided utterance;
 - (b) using the hint to establish a condition for distinguishing the provided utterance; and
 - (c) selecting a recognition result, derived in conjunction with operation of the speech recognizer, that satisfies the condition.
- 10 2. A method according to claim 1, wherein step (c) includes:
 - (i) providing a list of alternative recognition possibilities and
 - (ii) filtering the list based upon the condition.
- 15 3. A method according to claim 2, wherein the step of providing a list of alternative recognition possibilities includes using the speech recognizer to provide entries in the list.
4. A method according to claim 2, wherein the step of providing a list of
20 alternative recognition possibilities includes utilizing a linked dictionary to provide entries in the list.
5. A method according to claim 2, wherein the step of providing a list of alternative recognition possibilities includes both using the speech recognizer
25 and utilizing a linked dictionary to provide entries in the list.
6. A method according to claim 1, wherein the hint includes reference to the context of previous dictation to characterize the provided utterance.
- 30 7. A method according to claim 1, wherein the hint is a linguistic property characterizing the provided utterance.

8. A method according to claim 7, wherein the hint is an orthographic property characterizing the provided utterance.
9. A method according to claim 7, wherein the hint is a morphological property characterizing the provided utterance.
10. A method according to claim 7, wherein the hint is a semantic property characterizing the provided utterance.
11. A method according to claim 1, wherein the hint is a fractional spelling of the provided utterance.
12. A method according to claim 1, wherein the hint is a grammatical term characterizing the provided utterance.
13. A method according to claim 1, of utilizing a speech recognizer to distinguish a provided utterance from one or more similar-sounding utterances when a plurality of speaker-specified hints is provided, the method comprising:
- (a) identifying the hints and associating with them the provided utterance;
 - (b) using the hints to establish conditions for distinguishing the provided utterance; and
 - (c) selecting a recognition result, derived in conjunction with operation of the speech recognizer, that satisfies the conditions.
14. An improved speech recognition system wherein the improvement renders the system capable of distinguishing a provided utterance from one or more similar-sounding utterances when a speaker-identified hint is provided and the improvement comprises:
- (a) a hint recognizer which identifies the hint and associates with it the provided utterance;
 - (b) a condition specifier, coupled to the hint recognizer, which uses the hint to establish a condition for distinguishing the provided utterance; and

(c) a result selector, coupled to the condition specifier, which selects a recognition result that satisfies the condition.

15. A system according to claim 14, wherein the result selector includes a filter
5 operative on a list of alternative recognition possibilities.

16. A system according to claim 15, wherein the improvement further comprises a dictionary, to which the result selector is coupled, to provide entries in the list of alternative recognition possibilities.

10

17. A system according to claim 14, wherein the hint includes reference to the context of previous dictation to characterize the provided utterance.

18. A system according to claim 14, wherein the hint is a linguistic property
15 characterizing the provided utterance.

19. A system according to claim 18, wherein the hint is an orthographic property characterizing the provided utterance

20 20. A system according to claim 18, wherein the hint is a morphological property characterizing the provided utterance.

21. A system according to claim 18, wherein the hint is a semantic property characterizing the provided utterance.

25

22. A system according to claim 14, wherein the hint is a fractional spelling of the provided utterance.

23. A system according to claim 14, wherein the hint is a grammatical term
30 characterizing the provided utterance.

24. A system according to claim 14, wherein a plurality of speaker-identified hints is provided and the improvement comprises:

(a) a hint recognizer for identifying the hints and associating with them the provided utterance;

5 (b) a condition specifier, coupled to the hint recognizer, for using the hints to establish conditions for distinguishing the provided utterance; and

(c) a result selector, coupled to the condition specifier, for selecting a recognition result that satisfies the conditions.

10 25. An improved speech recognition system wherein the improvement renders the system capable of distinguishing a provided utterance from one or more similar-sounding utterances when a speaker-identified hint is provided and the improvement comprises:

(a) means for identifying the hint and associating with it the provided
15 utterance;

(b) means for using the hint to establish a condition for distinguishing the provided utterance; and

(c) means for selecting a recognition result, derived in conjunction with operation of the speech recognizer, that satisfies the condition.

20 [72062]

1/2

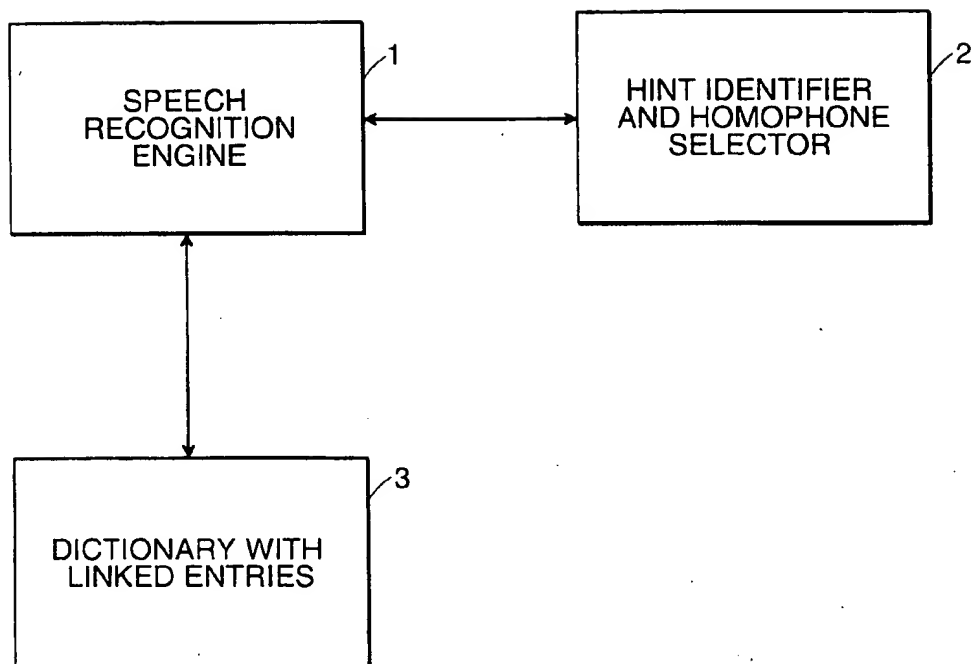


FIG. 1

2/2

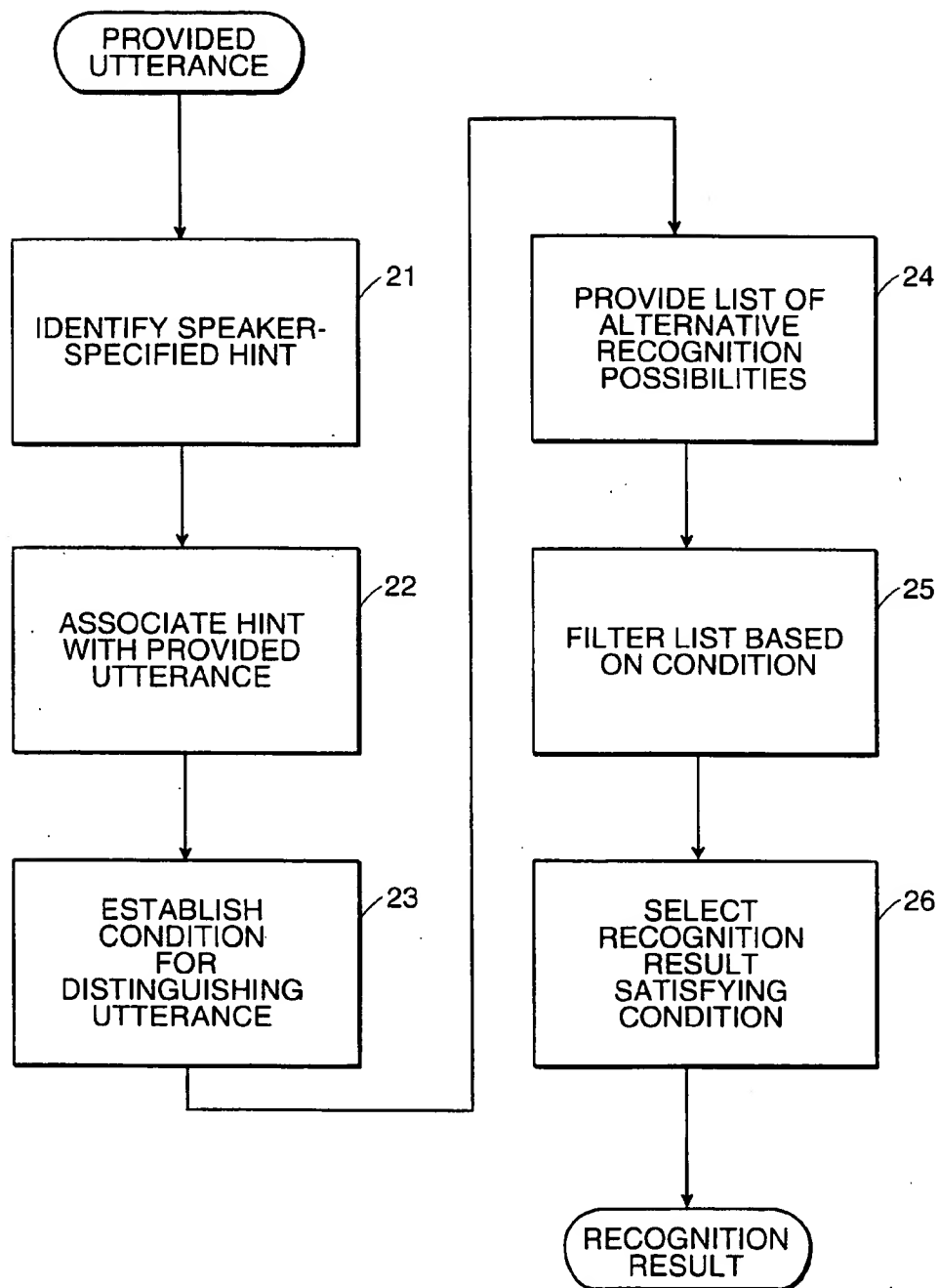


FIG. 2

INTERNATIONAL SEARCH REPORT

In national Application No

PCT/IB 98/01717

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 G10L5/06

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	FR 2 735 268 A (DRAGON SYSTEMS UK LTD) 13 December 1996 see page 3, line 24 - page 5, line 20 see page 6, line 22 - line 29. -----	1, 14, 25
A	US 5 054 074 A (BAKIS RAIMO) 1 October 1991 see column 3, line 38 - line 58 see column 6, line 18 - column 7, line 50 -----	1, 14, 25

☐ Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"S" document member of the same patent family

Date of the actual completion of the international search

14 January 1999

Date of mailing of the international search report

25/01/1999

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Krembel, L

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/IB 98/01717

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
FR 2735268 A	13-12-1996	CA 2178696 A	13-12-1996
US 5054074 A	01-10-1991	DE 69023365 D	14-12-1995
		EP 0385124 A	05-09-1990
		JP 2272498 A	07-11-1990